Solución a los ejercicios de autocomprobación - Tema 2

1. La información que nos dan es la siguiente:

$$n = 200,\ \ n_3 = 42,\ \ n_4 = 69,\ \ n_6 = 36,\ \ N_2 = 41,\ \ f_1 = 0.075$$

Además, el extremo superior de la primera clase es 75, el número de clases es 6 y la longitud de cada clase también es 6. Como no nos dicen cómo son los intervalos, los tomaremos cerrados en el extremo inferior y abiertos en el superior.

La tabla de frecuencias es la siguiente:

Intervalo	Marca	n_i	N_i	f_i	F_i
[69,75)	72	15	15	0.075	0.075
[75,81)	78	26	41	0.13	0.205
[81,87)	84	42	83	0.21	0.415
[87,93)	90	69	152	0.345	0.76
[93,99)	96	12	164	0.06	0.82
[99,105]	102	36	200	0.18	1
TOTAL		200		1	

Los valores que se piden son:

$\bar{x} = 88.35$, $\hat{\sigma}^2 = 75.93$, $\hat{\sigma} = 8.71$, $CV=0.098$, $M_e=88.478$, $P_{74}= 92.6521$

El intervalo modal es el [87, 93).

2. Los datos que necesitamos para el diagrama de caja son:

$Mín = 0.5$, $Máx = 4.5$, $Q_1 = 1.85$, $Q_2 = 2.361$, $Q_3 = 3.213$

El $R.I = Q_1 - Q_3 = 3.213 - 1.85 = 1.363.$

Las barreras quedarían:

$$Q_1 - 1.5\, R.I. = -0.1945, \qquad Q_3 + 1.5\, R.I. = 5.2575$$

Como la primera es más pequeña que el mínimo y la segunda mayor que el máximo, representamos el máximo y mínimo respectivamente en el diagrama. Marcamos $\bar{x} = 2.53$ con una cruz, a la derecha de la mediana.

3. Se tienen n=12 datos. Calculamos:

$$\bar{x} = 2.34, \qquad \hat{\sigma}^2 = 0.383, \quad \hat{\sigma} = 0.62$$

Calculamos los coeficientes de asimetría y curtosis:

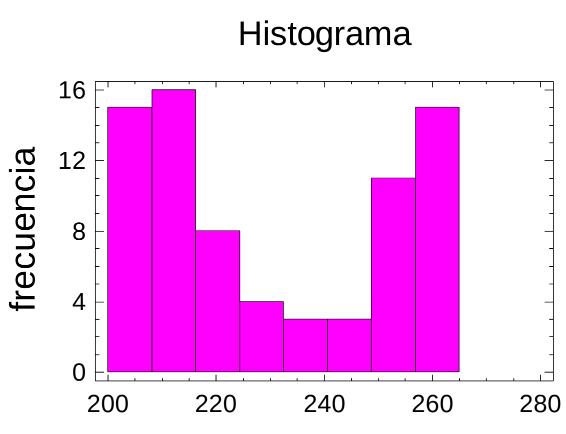
$$\gamma_1 = \frac{m_3}{\hat{\sigma}^3} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n\hat{\sigma}^3} = 0.035$$

$$\gamma_2 = \frac{m_4}{\hat{\sigma}^4} - 3 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n\hat{\sigma}^4} - 3 = -1.23$$

El coeficiente de asimetría es casi cero, con lo que la distribución es casi simétrica. Recordemos que la distribución es simétrica si este coeficiente es cero. La mediana, que es el punto medio entre los datos que ocupan la posición 6 y 7 en la serie de datos ordenada (2,41 y 2,45), es 2.43, un poquito por encima de la media.

El coeficiente de curtosis o apuntamiento es negativo con lo que la distribución es un poco más achatada que la distribución normal, que es la que se toma como referencia.

4. Como se representan los datos mensuales de venta en los últimos 12 años y medio, se representan un total de 78 datos. La venta máxima en un mes ha sido de 263 unidades y la mínima de 200 unidades. Vemos que la mediana es 220, menor que la media que es 230. En un rango de 20 unidades, del 200 (que es el mínimo) al 220, se sitúan la mitad de los datos y la otra mitad en un rango de 43, del 220 al máximo, que es 263. Además, las colas de la distribución no son iguales, la de la izquierda es mayor que la de la derecha, con lo que la distribución no es simétrica. Con todo esto podemos afirmar que la asimetría que presenta la distribución es positiva. No existen datos atípicos, todos se encuentran dentro de las barreras interiores. El rango intercuartílico (la longitud de la caja) es 43, grande comparado con el rango de la variable que es el máximo menos el mínimo, es decir, 63. Mientras más ancha sea la caja comparada con las barreras interiores, más forma de U tiene la distribución. A continuación se presenta un histograma al que podría corresponder el diagrama de caja que hemos comentado.



5. Algunos cálculos previos:

$$\sum_{i=1}^{8} x_i y_i = 6217, \quad \bar{x} = 5, \quad \bar{y} = 128, \quad \hat{\sigma}_X^2 = 5.5, \quad \hat{\sigma}_Y^2 = 3478.5, \quad Cov(X,Y) = 137.12$$

El valor del coeficiente de correlación es $\rho = 0.9936$. Este valor positivo y tan cercano a 1 indica que existe una relación lineal muy fuerte entre las dos variables. Conforme aumenta el número de semanas en el curso, la ganancia en rapidez de lectura aumenta de forma lineal con la primera. De hecho, el diagrama de dispersión para este conjunto de puntos nos muestra claramente esa relación lineal:

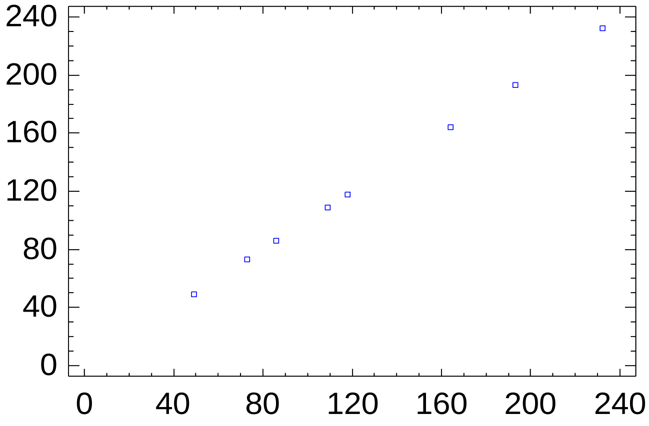


Diagrama de dispersión del ejercicio 5

6. i. Para calcular el gasto medio familiar, debemos calcular la media de la variable Y. Para ello, construimos primero la tabla de frecuencias de la distribución marginal de Y:

Clase	Marca	n_i	f_i	N_i	F_i
[0,150)	75	21	0.24	21	0.24
[150,300)	225	54	0.62	75	0.86
[300,450)	375	12	0.14	87	1
TOTAL		87	1		

$$\bar{y} = \frac{(75 * 21) + (225 * 54) + (375 * 12)}{87} = 209.48$$

Por lo tanto, el gasto medio en ocio por familia es de 209.38 euros.

ii. Para calcular los ingresos mensuales medios por familia, debemos calcular la media de la variable X. Para ello, construimos primero la tabla de frecuencias de la distribución marginal de X:

Clase	Marca	n_i	f_i	N_i	F_i
[600,1000)	800	8	0.09	8	0.09
[1000,1500)	1250	7	0.08	15	0.17
[1500,2000)	1750	38	0.44	53	0.61
[2000,2500)	2250	9	0.10	62	0.71
[2500,3000)	2750	25	0.29	87	1
TOTAL		87	1		

$$\bar{x} = \frac{(800 * 8) + (1250 * 7) + (1750 * 38) + (2250 * 9) + (2750 * 25)}{87} = 1961.49$$

Por tanto, el ingreso medio familiar es de 1961.49 euros.

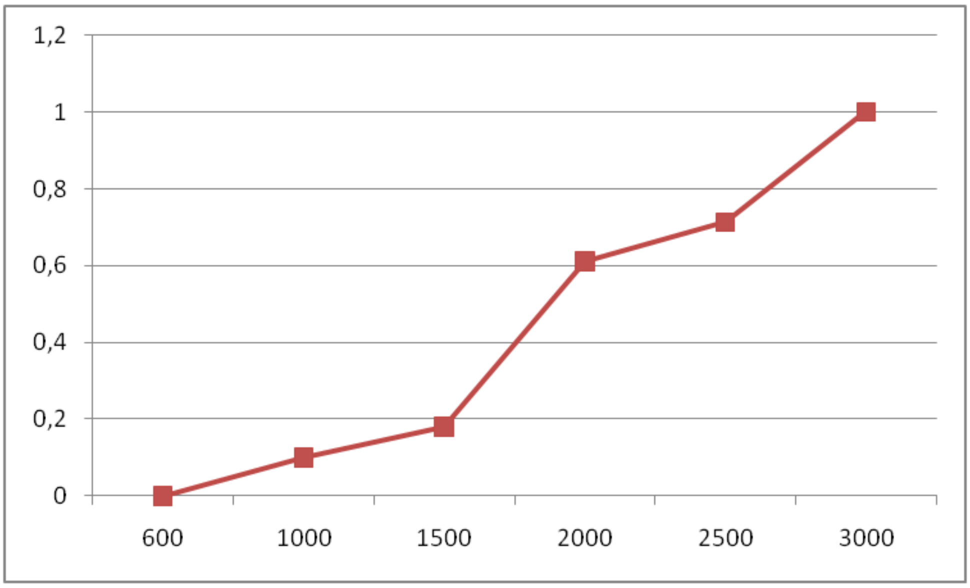
iii. Debemos calcular la distribución de $Y|X < 1500$. Para ello construimos su tabla de distribución de frecuencias:

Clase	Marca	n_i
[0,150)	75	9
[150,300)	225	5
[300,450)	375	1
TOTAL		15

$$\overline{y|x < 1500} = \frac{75 * 9 + 225 * 5 + 375 * 1}{15} = \frac{2175}{15} = 145$$

Por tanto, el gasto medio en ocio para familias con ingresos inferiores a 1500 euros es de 145 euros.

iv. Se pide el primer cuartil Q_1de la distribución de la variable $X = ingresos\ mensuales$. Para ello construimos el polígono de frecuencias acumuladas:



El intervalo en el que se encuentra es el [1500,2000). Interpolamos en él y obtenemos

$Q_1 = 1590.9$

Si aplicamos la fórmula correspondiente obtendríamos:

$$Q_1 = 1500 + \frac{1 * \frac{87}{4} - 15}{38} 500 = 1588.82$$

La pequeña diferencia entre ambos valores se debe al redondeo que debemos hacer en el cálculo de las frecuencias relativas en la distribución de *X*.

v. Para calcular la covarianza, utilizamos la fórmula:

$$Cov(X,Y) = \frac{\sum\sum n_{ij} x_i' y_i'}{n} - \bar{x}\bar{y}$$

donde x_i' e y_j' son las correspondientes marcas de clase. Se obtiene:

$$Cov(X,Y) = \frac{6 * 800 * 75 + 2 * 800 * 225 + \cdots + 2 * 2750 * 375}{87} - (1961.49) * (209.48)$$
$$= 17419.73$$

Calculamos ρ:

$$\rho = \frac{17419.73}{(91.07) * (609.58)} = 0.31$$

Su valor es pequeño con lo que la relación lineal entre ambas variables no es muy fuerte. Al ser positivo indicaría que la relación es directamente proporcional, es decir, conforme aumentan los ingresos mensuales aumentarían los gastos en ocio de las familias.

8. La media del conjunto transformado de datos es la media del primer conjunto de datos multiplicada por 4 más 3 unidades. La varianza del conjunto transformado de datos es la varianza del primero multiplicada por $16 = 4^2$. El sumar 3 no afecta al valor de la varianza en los datos transformados.

9. La primera afirmación se corresponde con el segundo gráfico. La segunda con el primer gráfico. La tercera afirmación se corresponde con el tercer gráfico.